1 Single-trial analysis of EEG during rapid visual discrimination: Enabling cortically-coupled computer vision

Paul Sajda, Adam D. Gerson, Marios G. Philiastides and Lucas C. Parra

1.1 Abstract

We describe our work using linear discrimination of multi-channel electroencephalography for single-trial detection of neural signatures of visual recognition events. We demonstrate the approach as a methodology for relating neural variability to response variability, describing studies for response accuracy and response latency during visual target detection. We then show how the approach can be utilized to construct a novel type of brain-computer interface, which we term cortically-coupled computer vision. In this application, a large database of images is triaged using the detected neural signatures. We show how 'cortical-triaging' improves image search over a strictly behavioral response.

1.2 Introduction

Running in the park with your head phones on, listening to your favorite tune and concentrating on your stride, you look up and see a face that you immediately recognize as a high school friend. She is wearing a hat, glasses, and has aged 15 years since you last saw her. You and she are running in opposite directions so you only see her for a fleeting moment, yet you are sure it was her. Your visual system has just effortlessly accomplished a feat that has thus far baffled the best computer vision systems. Such ability for rapid processing of visual information is even more impressive in light of the fact that neurons are relatively slow processing elements compared to digital computers, where individual transistors can switch a million times faster than a neuron can spike.

Non-invasive neuroimaging has provided a means to peer into the brain during rapid visual object recognition. In particular, analysis of trial-averaged event-related potentials (ERPs) in electroencephalography (EEG), has enabled one to assess the speed of visual recognition and discrimination in terms of the timing of the underlying neural processes [33]. More recent work has used single-trial analysis of EEG to characterize the neural activity directly correlated with behavioral variability during tasks involving rapid visual discrimination [7, 28]. These results suggest that components extracted from the EEG can

capture the neural correlates of the visual recognition and decision making processing on a trial-by-trial basis.

In this paper we consider how such EEG components might be utilized for constructing a brain computer interface (BCI) system for rapidly assessing streams of natural images. Traditionally, non-invasive BCI systems have been based on one of the following paradigms; 1) having a subject consciously modulate brain rhythms (e.g. [26, 36, 6]), 2) having a subject consciously generate a motor plan and/or visual imagery [27, 35], 3) directly modulate the subject's cortical activity by the stimulus frequency (e.g. steady state visually evoked potentials SSVEP) [16, 4] or 4) exploit specific ERPs such as the novelty/oddball P300 [15]. The approach and system we describe is most similar to the later, though our focus is on single-trial detection of ERPs and their relationship to visual discrimination and recognition.

We begin this paper by providing a brief review of the linear discrimination methods we employ to extract task specific components in the EEG. We then show how such components are in fact directly coupled with the visual discrimination and decision making processes for stimuli involving rapid sequences of natural images. For example, we show that we can construct neurometric functions from the EEG components which are indistinguishable from the corresponding psychometric functions for a rapid serial visual presentation (RSVP) task. We also investigate the neural correlates of response time variability responsible for such perceptual decision making processes. We then describe how we use this approach to develop a BCI system high-throughput imagery triage. We term our system *cortically-coupled computer vision* since we leverage the robust recognition capabilities of the human visual system (e.g. invariance to pose, lighting, scale, etc), and use a noninvasive cortical interface to intercept signatures of recognition events–i.e. the visual processor performs perception and recognition and the EEG interface detects the result (decision) of that processing.

1.3 Linear methods for single-trial analysis

The goal of a BCI system is to detect neuronal activity associated with perceptual and/or cognitive events. Detecting such events implies detecting when an event occurred and identifying its significance. The task is greatly simplified if the timing information is provided by an external observable event. Thus the conventional paradigm of the evoked response considers the neuronal activity following the presentation of a stimulus. In our work we have adopted this paradigm by analyzing the EEG activity of multiple electrodes following presentation of an image. For simplicity we aim to identify only one type of event, visual target recognition, and differentiate this from other visual processing. The task is therefore a binary classification based on the temporal and spatial profile of the potentials evoked following stimulus presentation. In every trial an image is presented and in some trials the image contains a target object which we assume is recognized by the subject. The EEG activity following each stimulus is recorded as $D \times T$ values, where D is the number of channels and T is the number of samples. Typically we record data at 1000 Hz in up to 64 channels. With a time window of half a second following the

presentation of the stimulus, one would have acquired 32000 samples. This is a rather large feature vector considering that typically there are fewer than $N = 100$ exemplars (trials) to train a classifier. In addition, EEG signals have a very low signal-to-noise ratio (SNR) and brute-force classification of these 32000-dimensional feature vector will typically fail.

To obtain reasonable classification performance we exploit prior information on the temporal characteristics of the signal and noise with the following steps: (1) Reduce the trial-to-trial variability by filtering the signal to remove 60Hz interference and slow drifts (slower than 0.5 Hz). This assumes that slow constant currents below 0.5 Hz carry no information; (2) Reduce the dimensionality of the problem by stepping our classification window every L-th sample assuming that the signal of interest does not vary much within L samples; (3) Increase the number of exemplars by using the L redundant samples in each classification window. This implies that the variation within L samples is considered noise, i.e. for $L = 50$ the signal of interest is at 10Hz while faster signal variation are considered noise. Steps (2) and (3) taken together will transform the original data for each trial with TD dimensions into L exemplars of only DT/L dimensions. As an example, with $L = 50$ and $N = 100$ one will acquire 5000 training examples, which can be used to train a classifier with a 640-dimensional feature vector. Admittedly this samples are not independent, but they are useful as they capture the noise in the data at least for frequencies above 10Hz.

We have obtained good classification results with a simple linear classifier of these DT/L-dimensional feature vectors. The classification method is demonstrated in Figure 1.1 for the simple case of a single training window ($L = T$ and D-dimensional feature vector). Linear classification means that the feature vector $\mathbf{x}$ is projected onto an orientation defined by vector $\mathbf{w}$ such that the projection, $y = \mathbf{w}^T\mathbf{x}$, optimally differentiates between the two classes. This is a traditional problem in pattern recognition with various solutions depending on the exact optimality criteria. In an off-line processing mode we use penalized logistic regression as it gives us the best generalization performance on this data [25]. For well-separated classes this linear classification method is equivalent to linear support vectors. In an real-time processing mode we use Fisher linear discriminants as the required means and covariances can easily be updated on-line as more trials become available for training. For a discussion on the relative benefits of various linear classification methods with EEG data see [23, 25]. Classification performance is measured with the conventional receiver operating characteristic (ROC) curve [8], specifically the area under the ROC curve (Az). We will report in all cases the cross-validated test-set performance using a leave-one-out procedure where we leave out all samples belonging to one trial.

One can conceive of many other ways of classifying the spatio-temporal evoked responses including non-linear methods. In fact, many different algorithms have been proposed, which exploit different prior assumptions on the signals [25, 18, 20, 3]. We are partial towards linear methods for two reasons: (1) The linear combination of voltages has an immediate interpretation as a current (tissue is primarily resistive with coupling coefficients representing conductivity). The coefficients that couple this current with the observed voltages are given for the linear model by, $\mathbf{a} = \langle \mathbf{x}^T y \rangle / \langle y^2 \rangle$, where the angular brackets indicate the average over trials and samples. Specifically, coefficients $\mathbf{a}$ describe the coupling (and correlation) of the discriminating component y with the sensor activity

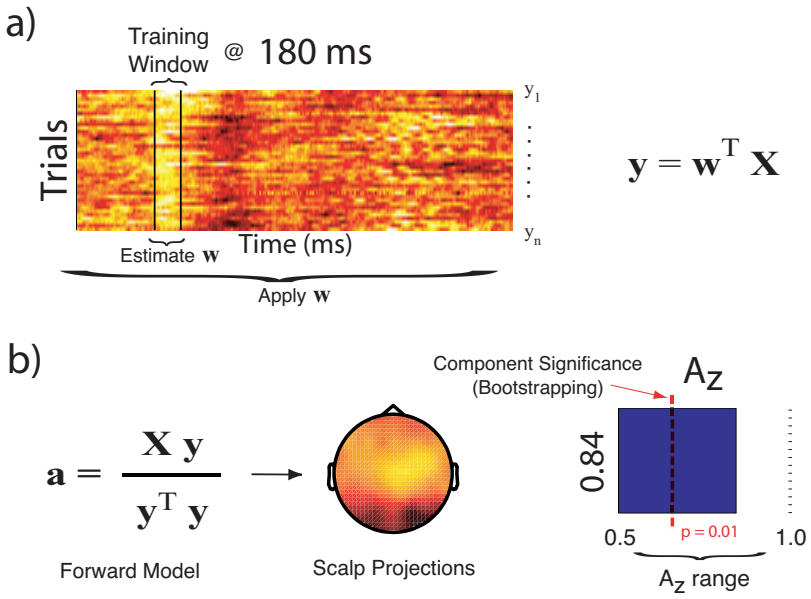


Figure 1.1: Linear discrimination in EEG: (a) The D-dimensional EEG activity, $\mathbf{X}$, is projected onto a single dimension $\mathbf{y}$. ($\mathbf{X}$ is a matrix of channels by samples, and $\mathbf{y}$ is a row vector containing multiple samples). The row vectors, $\mathbf{y}$, containing the samples that follow each target stimulus presentation, can be arranged for multiple trials as a matrix. This matrix ($\mathbf{y_{target}}$ - mean($\mathbf{y_{non-target}}$)) is displayed here as an image with white and black representing the largest and smallest values respectively. The projection vector $\mathbf{w}$ is chosen so that the values $\mathbf{y}$ within the training window differ maximally between target and non-target trials. (b) The sensor projections $\mathbf{a}$ are computed for the samples within the training window. (In this equation the inner product computes the average over trials and samples. Therefore matrix $\mathbf{X}$ and vector $\mathbf{y}$ extend here over the training samples from all trials.) The resulting values of $\mathbf{a}$ are displayed at the corresponding scalp locations as a color-map with white and black representing the largest and smallest values respectively. When the intensity $\mathbf{y}$ averaged within the specified time window is used as classification criteria we achieve on this data an Az-value of 0.84. The probability of obtaining an Az of this magnitude by chance is less than 1% ($p < 0.01$).

$\mathbf{x}$. Both $\mathbf{a}$ and $\mathbf{x}$ are D-dimensional vectors (row and column respectively). Strong coupling indicates low attenuation of the component and can be visualized as intensity maps that we call the 'sensor projections' [25], (2) Linear methods are easy to implement and are fast, permitting real-time operation. The disadvantage of our method is that it does not capture synchronized activity above 10Hz, and neither does it capture activity that is not at a fixed distance in time from the stimulus, instead only phase-locked activity is detected.

In the remaining sections we give several examples of how this linear discrimination method is used to identify the neural correlates of decision making and response time variability, as well as how it can be integrated into a BCI system for image triage.

1.4 EEG correlates of perceptual decision making

Identifying neural activity directly responsible for perceptual decision making is a major challenge for non-invasive BCI systems. A number of investigators have studied the neural correlates of decision making in awake behaving animals, in particular primates, where single and multi-unit recordings have been analyzed using signal detection theory [8] and subsequently correlated with the animal's observed behavior [2, 1, 24]. These approaches consist mainly of direct comparisons between psychometric and neurometric functions since this enables one to relate the variability of the neural activity to the variability observed in the behavioral response. The technique has been applied in a variety of perceptual decision making paradigms, including discrimination of visual objects such as faces [17]. The approach, though powerful, has been limited to animal studies which use invasive recordings of single-trial neural activities. Yet to be demonstrated however is whether decision making could be studied in a similar fashion, non-invasively, in humans.

We use single-trial linear discrimination analysis, as outlined in the previous section, to identify the cortical correlates of decision making during rapid discrimination of images. Psychophysical performance is measured for several subjects during an RSVP task, where a series of target (faces) and non-target (cars) trials are presented in rapid succession (Fig. 1.2a), while simultaneously recording neuronal activity from a 64-channel EEG electrode array. Stimulus evidence is varied by manipulating the phase coherence [5] of the images (Fig. 1.2b). Within a block of trials, face and car images over a range of phase coherences are presented in random order. We use a set of 12 face (Max Planck Institute face database) and 12 car grayscale images (image size 512 x 512 pixels, 8-bits/pixel). Both image types contained equal numbers of frontal and side views (up to 45 degrees). All images are equated for spatial frequency, luminance and contrast. Subjects are required to discriminate the type of image (face or car) and report their decision by pressing a button.

EEG data is acquired simultaneously in an electrostatically shielded room (ETS-Lindgren, Glendale Heights, IL) using a Sensorium EPA-6 Electrophysiological Amplifier (Charlotte, VT) from 60 Ag/AgCl scalp electrodes and from three periocular electrodes placed below the left eye and at the left and right outer canthi. All channels are referenced to the left mastoid with input impedance $< 15k\Omega$ and chin ground. Data are sampled at 1000 Hz with an analog pass band of 0.01–300 Hz using 12 dB/octave high pass and eighth-order elliptic low pass filters. Subsequently, a software based 0.5 Hz high pass filter is used to remove DC drifts and 60 and 120 Hz (harmonic) notch filters are applied to minimize line noise artifacts. These filters are designed to be linear-phase to minimize delay distortions. In all our experiments we record also EOG signals and remove motion and blink artifacts using linear methods as described in [25]. Motor response and stimulus events recorded on separate channels are delayed to match latencies introduced by digitally filtering the EEG.

Using a linear discriminator, we identify EEG components that maximally discriminate between the two experimental conditions. At each phase coherence level, and between the stimulus onset and the earliest reaction time, we identify two time windows which gave the most discriminating components. For this paradigm an early ($\approx$ 170 ms following

stimulus) and a late component (> 300 ms following stimulus) can be identified. In order to be able to directly compare the neuronal performance at these two times, to the psychophysical sensitivity as captured by the psychometric functions [8], we construct neurometric functions by plotting the area under the ROC curves (Az values) against the corresponding phase coherence levels. A linear discriminator is trained by integrating data across both time windows ($2D$-dimensional feature vector). With this approach, we generally observe for the discriminator improved performance (and hence higher Az values) compared to when training is performed on the individual components in isolation. Figure 1.3 shows a comparison of the psychometric and neurometric functions for one subjects in the dataset. To demonstrate that the EEG-derived neurometric functions can account for psychophysical performance, a likelihood-ratio test is used [11] which shows that for all the subjects a single function can fit the behavioral and neuronal data sets as well as the two separate functions.

For both the early (the well-known N170 [31, 29, 12]) and late face selective components, at each phase coherence level, we construct discriminant component maps to help us visualize the temporal evolution of the discriminating activity across trials. Data is analyzed for both stimulus and response-locked conditions, showing that both face selective components appear to be more correlated with the onset of visual stimulation rather than the response as shown in Figure 1.4 for one subject. In addition we construct scalp maps of these discriminating components. The spatial distribution of activity seems to indicate signaling between occipito-parietal and centro-frontal networks, consistent with several ERP/MEG and functional neuroimaging studies [9, 10, 19, 31, 34]. The Az values which describe the discriminator's performance at each phase coherence level are also shown. For the subject shown in Figure 1.4, the discriminant activity is statistically significant down to a 30% phase coherence for both the early and late components as assessed by a bootstrapping technique. Specifically, we compute a significance level for Az by performing the leave-one-out test after randomizing the truth labels of our face and car trials. We repeat this randomization process 100 times to produce an Az randomization distribution and compute the Az leading to a significance level of $p < 0.01$.

Our results demonstrate that neural correlates of perceptual decision making can be identified using high-spatial density EEG and that the corresponding component activities are temporally distributed. Clearly important to the identification of these neural correlates is the spatial, and to a lesser extent the temporal integration of the EEG component activities. This approach is complementary to approaches using single and multi-unit recordings since it sacrifices spatial and some temporal resolution (local field potentials versus spike-trains) for a more spatially distributed view of the neural activity during decision making. The fact that we are able to identify neural correlates of perceptual decision making using relatively poor spatial resolution of EEG suggests that these neural correlates represent strong activities of neural populations and not the activity of a small number of neurons. As such, this approach can be proven especially useful in designing non-invasive BCI systems that can reliably predict behavioral responses.

1.5 Identifying cortical processes leading to response time variability

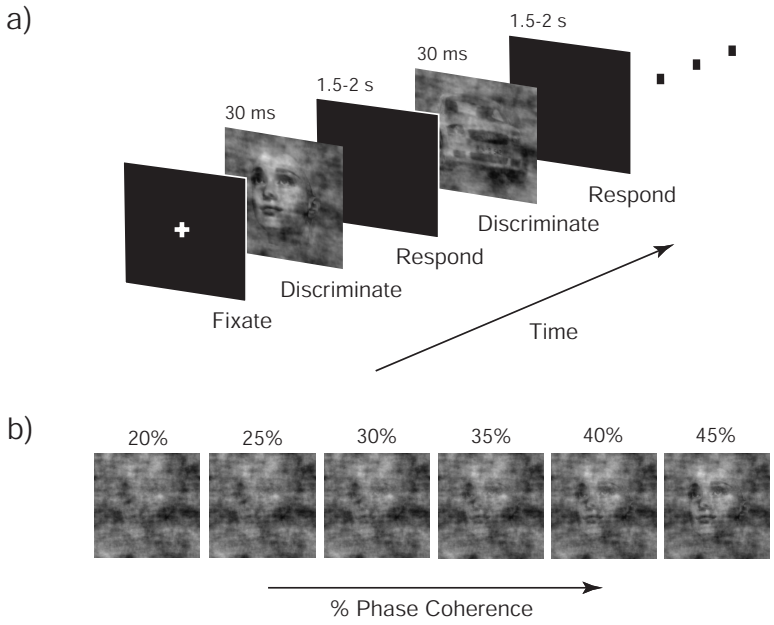


Figure 1.2: Schematic representation of the behavioral paradigm. (a) Within a block of trials subjects are instructed to fixate on the center of the screen and are subsequently presented, in random order, with a series of different face and car images at one of the six phase coherence levels shown in (b). Each image is presented for 30 ms followed by an inter-stimulus-interval lasting between 1500-2000 ms during which subjects are required to discriminate among the two types of images and respond by pressing a button. A block of trials is completed once all face and car images at all six phase coherence levels are presented. (b) A sample face image at 6 different phase coherence levels (20, 25, 30, 35, 40, 45%).

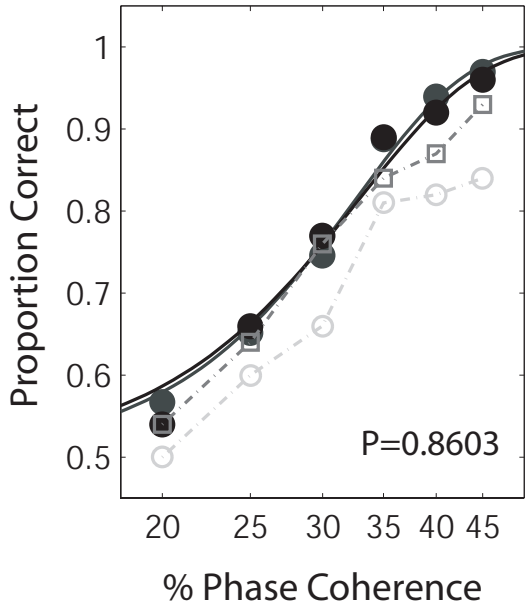


Figure 1.3: Comparison of behavioral and neuronal performance. Psychometric (solid gray) and neurometric (solid black) functions for one subject. The abscissas represent the percentage of phase coherence of our stimuli and the ordinate indicates the subject's performance as proportion correct. We fit both data with separate Weibull functions [30]. The psychophysical and neuronal data are statistically indistinguishable as assessed by a likelihood ratio test after we fit the best single Weibull function jointly to the two data sets. The *p*-value in the bottom right corner represents the output of this test. A *p*-value greater than 0.05 indicates that a single function fits the two data sets as well as the two separate functions. The dotted gray lines connect the A_z values computed for each of the two training windows separately (earlier window, light gray circles; later window, dark gray squares).

Significant variability in response time is observed across trials in many visual discrimination and recognition tasks. A variety of factors may account for response time variability ranging from the difficulty in discriminating an object on any given trial, trial-by-trial variability of the subject's engagement in the task, or intrinsic variability of neural processing. Identifying neural activity that is correlated with response time variability may shed light on the underlying cortical networks responsible for perceptual decision making processes and the processing latencies that these networks may introduce for a given task.

We study visual target detection using an RSVP paradigm and use single-trial spatial integration of high-density electroencephalography to identify the time course and cortical origins leading to response time variability. The RSVP task emulates natural saccadic scene acquisition and requires high vigilance. The RSVP paradigm is illustrated in Figure 1.5.

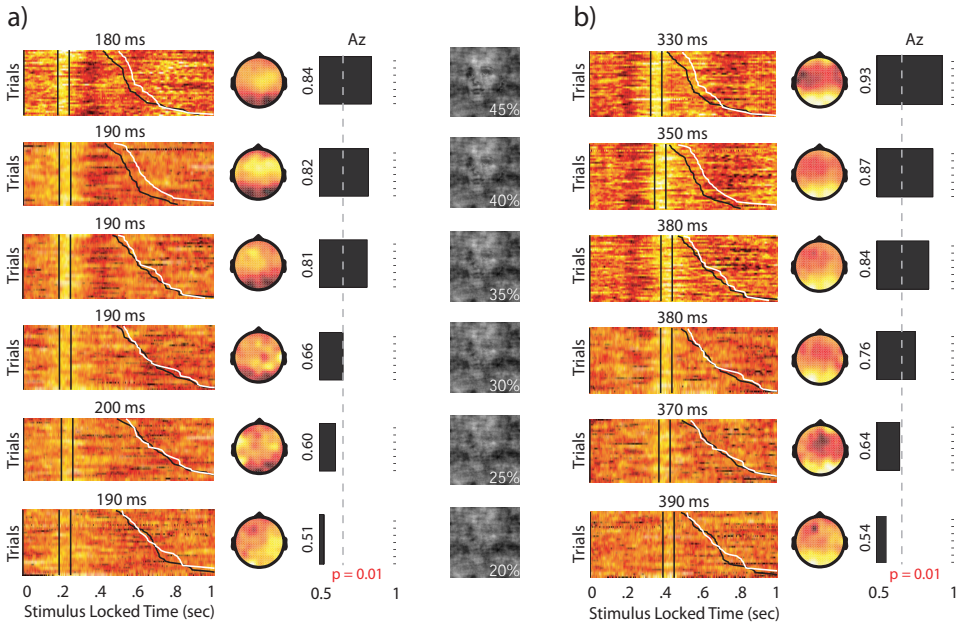


Figure 1.4: Discriminant component activity that shows the difference between face vs. car trials at each coherence level for one subject for (a) the early (N170) and (b) the late ($\approx 300-400$ ms) window. White represents positive and black negative activity. Each row of these maps represents the output of the linear discriminator for a single trial, using a 60 ms training window (vertical white lines) with onset times specified at the top of each panel. All trials are aligned to the onset of visual stimulation, as indicated by the vertical black line at time 0 ms, and sorted by response time. The black and white sigmoidal curves represent the subject's response times for face and car trials respectively. The representation of the topology of the discriminating activity is shown by the scalp plots to the right (dorsal view). White represents positive correlation of the sensor readings to the extracted activity and black negative correlation. The A_z values for each time window at each coherence level are represented by the bar graphs. The significance of the difference activity is represented by the dotted line ($p = 0.01$).

Activity associated with recognition has been identified with the RSVP paradigm as early as 150ms after stimulus presentation [33]. More recent work argues that this activity is associated with differences in low level features of the imagery rather than target recognition [13]. The varied scale, pose and position of target objects (people) requires subjects to recognize objects rather than low level features. During this task, participants are presented with a continuous sequence of natural scenes. Participants completed four blocks of 50 sequences each with a rest period lasting no more than five minutes between blocks. Each sequence consists of 50 images and have a 50% chance of containing one target image with one or more people in a natural scene. These target images can only appear within the middle 30 images of each 50 image sequence. The remaining natural

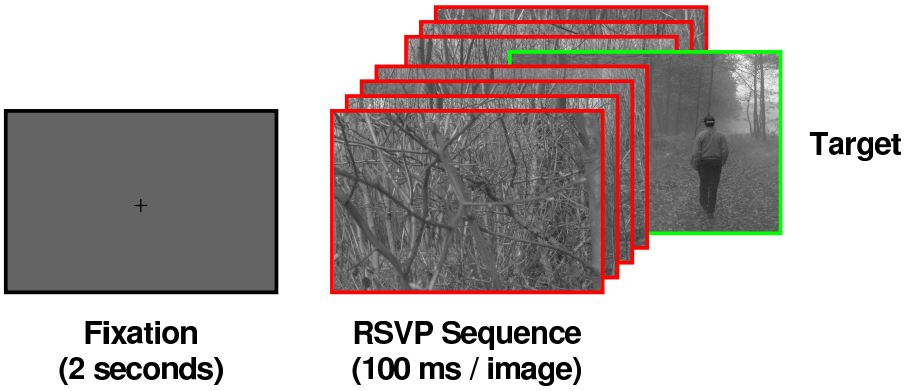


Figure 1.5: Example Rapid Serial Visual Presentation (RSVP) trial. A fixation cross lasting two seconds is followed by a sequence of 50 images. Each sequence has a 50% probability of containing one target image. This target can only appear within the middle 30 images to ensure that a one second image buffer precedes and follows the target.

scenes without a person are referred to as distractor images. Each image was presented for 100 ms. A fixation cross is displayed for 2 seconds between sequences. Participants are instructed to press the left button of a generic 3-button mouse with their right index finger while the fixation cross is present, and release the button as soon as they recognize a target image.

Linear discrimination is used to determine spatial weighting coefficients that optimally discriminate between EEG resulting from different RSVP task conditions (e.g. target vs. distractor images) over specific temporal windows between stimulus and response. Integration across sensors enhances signal quality without loss of temporal precision common to trial averaging in ERP studies. The resulting discriminating components describe activity specific to target recognition and subsequent response for individual trials.

Inter-trial variability is estimated by extracting features from discriminating components. While robust extraction of component onset from individual trials is extremely difficult due to the stochastic nature of EEG, there is evidence of strong correlation between ERP peak and onset times [32]. The peaks of spatially integrated discriminating components were found by fitting a parametric function to the extracted component $y(t)$. For simplicity we use a Gaussian profile that is parameterized by its height β, width σ, delay μ, and baseline offset α:

$$\hat{y}(t) = \alpha + \frac{\beta}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}. \tag{1.1}$$

Response locking of discriminating components is determined by computing the linear regression coefficients that predict the latency of the component activity as measured by μ from the response times given by r as described by Equation 1.2. The slope from the response time peak latency regression (θ) is defined to be the degree of response locking

(percentage) for each component. This metric quantifies the extent to which the component is correlated with the response across trials. It ranges from 0% for pure stimulus lock to 100% for pure response lock. A slope $\theta = 100\%$ indicates that slow responses show a corresponding late activity, and fast responses show a corresponding early activity. A slope of $\theta = 0\%$ indicates that the timing of the activity does not change with response time and is therefore stimulus locked.

$$\hat{\mu}_j = \theta r_j + b \tag{1.2}$$

where $\hat{\mu}_j$ and r_j are the predicted peak latencies and response times for the j–th trial and b is an offset term for the regression. This is shown for one subject in Figure 1.6.

The group results for the discriminating component activity across nine participants is shown in Figure 1.7. Scalp projections of discriminating components were normalized prior to averaging. Group averaged results show a shift of activity from frontal to parietal regions over the course of 200 ms, which is consistent with previous studies of visual oddball [22, 21]. Additional analysis and discussion is provided in [7].

In order to estimate the progression of response locking across all subjects, it is necessary to account for response time variability between subjects. It is not appropriate to average results since components are not temporally aligned across subjects. Rather, histograms of response times were equalized to one subject (subject 2), and component peak times were scaled accordingly. Scaled response times and component peak times were concatenated across subjects. These registered group response times were then projected onto the scaled component peak times to estimate the degree of response locking across subjects. The group response lock increases from 28% at -200 ms to 78% 50 ms after the response.

The features of discriminating components are believed to reflect visual processing, attention and decision stages. Modeling the peak latency, amplitude and duration of each trial allows us to study the extent to which each stage varies with response time. Consistent with [14], Figure 1.7 indicates that significant processing delays may be introduced by early processing stages. Within 200 ms prior to response (≈ 250 ms following stimulus), activity is already, on average, between 25-35% response locked. Due to our method, it is not possible to determine whether this response locking is a result of components at this onset time or earlier onset times, since discriminating components were not significant for earlier onset (peak) times. Thus we conclude it is possible that some of this early response locking may be due to early visual processes (0-250 ms poststimulus). For our nine subjects, correlation analysis reveals that discriminating component activity progressively becomes more response locked with subsequent processing stages. Along with scalp projections derived from discriminant analysis, the covariability of peak latency with response time describes which cortical regions introduce processing delays, providing insight into the nature of information flow through the brain during visual discrimination.

1.6 EEG–based image triage

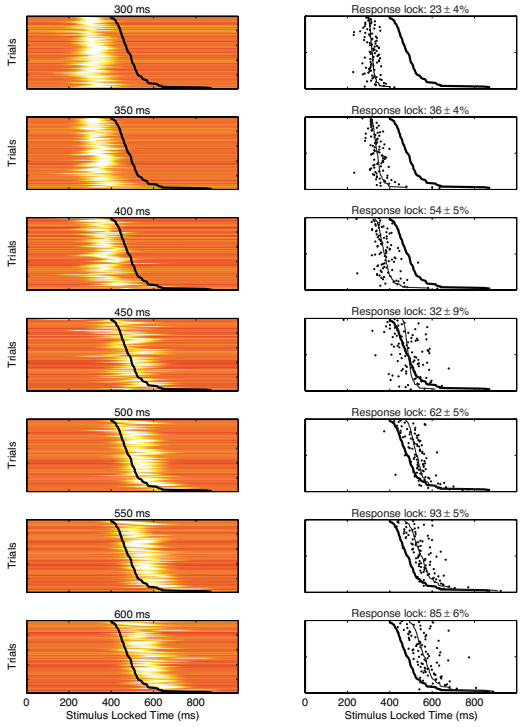


Figure 1.6: Detailed temporal analysis of stimulus locked discriminating activity for subject 2. Each row in the left column shows the fit of discriminating activity to a Gaussian profile described by Equation 1.1. On the top of each of these panels is the onset time of the window used for discrimination. Right columns of each panel display the peak latency (μ) (black dots) of each trial. The projection of response times onto these peak latencies is shown with a line black curve, with thick black curves representing response times. The parameters for this projection indicate the degree of response locking for each component. Purely stimulus and response locked conditions are indicated by 0% and 100% response lock respectively. On top of these panels are reported the percent response lock and corresponding error in the fit of the peak latencies across trials as well as the mean onset time of the component. The standard deviation of peak latencies is 62 ms.

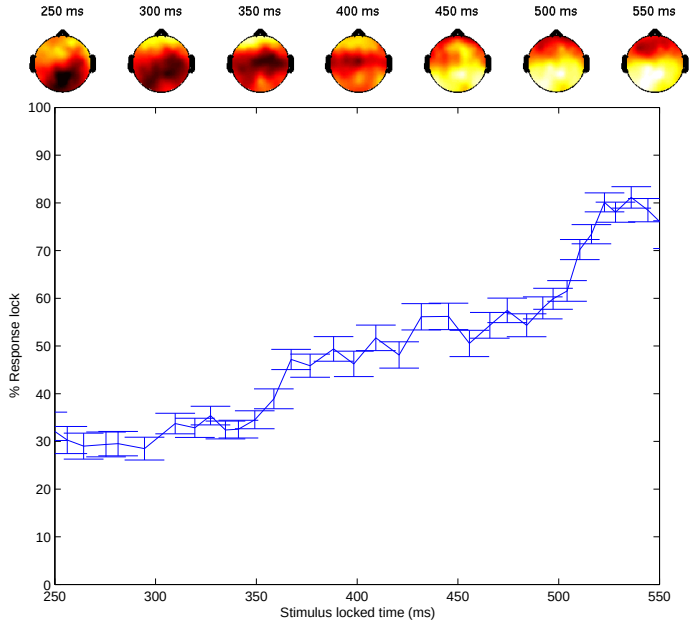


Figure 1.7: Group results over all 9 subjects for stimulus locked discriminating components. Top row shows scalp distribution of discriminating activity averaged over all subjects. Bottom row shows the degree of response locking over time. Error bars reflect standard error of the regression parameter associated with response locking %. For all subjects the first discriminating activity is frontal and correlated more with the stimulus than response. By the time it arrives in parietal areas a delay has been introduced.

Finally, we describe an EEG system capable of using neural signatures detected during RSVP to triage sequences of images, reordering them so that target images are placed near the beginning of the sequence. We term our system “cortically–coupled computer vision” since we leverage the robust recognition capabilities of the human visual system (e.g. invariance to pose, lighting, scale, etc), and use a non-invasive cortical interface (e.g. EEG) to intercept signatures of recognition events – the visual processor performs perception and recognition and the EEG interface detects the result (e.g. decision) of that processing.

The RSVP triage task is similar to the task described in Figure 1.5 however following the image sequence a series of self–paced feedback slides were presented indicating the position of target images within the sequence before and after EEG–based triage. Participants completed two blocks of 50 sequences with a brief rest period lasting no more than five minutes between blocks. During the second block, participants were instructed to quickly press the left button of a generic 3-button mouse with their right index finger as soon as they recognized target images. They were instructed to press the button twice, as quickly as possible, if one target image immediately followed the other. Participants did not respond with a button press during the first block.

In order to classify EEG on–line we use a Fisher linear discriminator to estimate a spatial weighting vector which maximally discriminates between sensor array signals evoked by target and non–target images. During each experimental condition (with and without motor response), 5000 images were presented to the subject in sequences of 100 images. EEG evoked by the first 2500 images (50 targets, 2450 non-targets) was used to train the classifier. During the experimental sessions, a training window between 400-500 ms following stimulus onset was used to extract training data. Weights were updated adaptively with each trial during the training period. Classification threshold is adjusted to give optimum performance for the observed prevalence (class-prior). These weights and threshold were fixed at the end of the training period and applied to the subsequent testing dataset (images 2501-5000).

To boost triage performance, after the experiment multiple classifiers with different training window onsets were used. The training window onsets ranged from 0 to 900 ms in steps of 50 ms. The duration of the training windows was 50 ms. Once these classifiers were trained, the optimal weighting of these classifier outputs was found using logistic regression to discriminate between target and non–target images.

Again, only EEG data evoked by the first 2500 images was used to train the classifiers and then find the inter–classifier weights. These weights were then applied to the testing data set evoked by the second set of 2500 images (images 2501-5000).

Following the experiment, all image sequences were concatenated to create training and testing sequences that each contain 2500 images (50 targets and 2450 non–targets). These image sequences are re–sorted according to the output of our classifier with multiple training windows for EEG evoked by every image.

For comparison, sequences were triaged based on the button response. Images were resorted according to:

$$p(\text{target}|RT) = \frac{p(RT|\text{target})p(\text{target})}{p(RT|\text{target})p(\text{target}) + p(RT|\text{non–target})p(\text{non–target})} \tag{1.3}$$

where RT is the onset of a button response that occurs within one second of image onset. $p(\text{target}|RT) = 0$ when no response occurred within one second of image onset. The priors $p(\text{target}) = 0.02$ and $p(\text{non–target}) = 0.98$. $p(RT|\text{target})$ is a Gaussian distribution with a mean and variance determined from the response times from the training sequences. Since more than one response is likely to follow a target image if the two target images are presented within one second of each other, for training sequences response times were assigned to target images based on the position of the target image within the sequence. In other words if the target appeared first in the sequence and two button responses occurred within one second of this target's onset, the first response was assigned to that target image and the second response was assigned to the second target image. For testing sequences, if two or more responses occur within one second of the onset of any image, the response with the greatest $p(\text{target}|RT)$ is assigned to the image. $p(RT|\text{non–target})$ is a mixture of 13 Gaussians, each with the same variance as that used for $p(RT|\text{target})$ and with means assigned by shifting the mean from $p(RT|\text{target})$ 600 ms in the past to 700 ms in the future in increments of 100 ms, excluding the actual mean of $p(RT|\text{target})$. This mixture model contains a sufficient number of Gaussians so that the mixture is consistent within the one second interval following image onset. $p(RT|\text{non–target})$ was designed to model responses occurring within one second of the onset of a non–target image that is presented within one second prior to or following a target image.

Triage results for one subject (subject 2) are shown in Figure 1.8. Figure 1.8(a) shows number of targets as a function of the number of distractor images both before and after triage based on button press and EEG. The area under the curve generated by plotting fraction of targets as a function of the fraction of distractor images presented is used to quantify triage performance. Triage performance for five subjects is listed in Table 1.1. This area is 0.50 for all unsorted image sequences since target images are randomly distributed throughout the sequences. Ideal triage performance results in an area of 1.00. There is no significant difference in performance between button–based and EEG–based triage ($0.93 \pm 0.06, 0.92 \pm 0.03, p = 0.69, N = 5$). Interestingly there is no significant difference in performance between EEG–based triage for the motor and no motor response conditions ($0.92 \pm 0.03, 0.91 \pm 0.02, p = 0.81, N = 5$).

Figures 1.8(b)–1.8(f) are rasters showing the position of the target images (black squares) and non–target images (white squares) in the concatenated image sequence. Based on these rasters and the EEG and button–based triage performance for five subjects list in Table 1.1, it is clear that both EEG and button–based triage systems are capable of a high level of performance. The button–based triage performance begins to fail, however, when subjects do not consistently respond to target stimuli and response times exceed one second. Subject 2, for instance correctly responded to only 74% of targets during the testing session. In fact, this subject did not respond to 12 of 50 target images and the response time for 1 target image exceeded one second. Excessively late responses cannot effectively

Table 1.1: Triage performance and behavioral results

Subject	EEG (no motor)	EEG (motor)	Button	EEG (motor) and Button	RT (training) (ms)	RT (testing) (ms)	% Correct (training)	% Correct (testing)
1	0.92	0.91	0.87	0.94	418 ± 133	413 ± 101	88	86
2	0.94	0.96	0.86	0.97	412 ± 64	450 ± 64	94	74
3	0.90	0.87	0.96	0.96	445 ± 79	423 ± 59	86	94
4	0.91	0.92	0.98	0.98	433 ± 74	445 ± 59	98	98
5	0.91	0.93	0.98	0.98	398 ± 86	402 ± 58	96	96
Group	0.91 ± 0.02	0.92 ± 0.03	0.93 ± 0.06	0.97 ± 0.02	421 ± 91	426 ± 71	92 ± 5	90 ± 10

be classified using our bayesian methods since it is not clear whether these button presses were in response to the target image or a subsequent non–target image. The EEG response evoked by images with either no response or a late response is, however, still consistent with EEG evoked by the target images with predictable response times. The EEG–based triage system is therefore capable of detecting the recognition of these target images and subsequently resorting these target images appropriately. For this reason we exploit the information provided by both EEG and button press using another perceptron to boost triage performance. This approach is effective for increasing triage performance for subjects that either did not respond or had a delayed motor response to a significant number of target images (e.g. subjects 1 and 2).

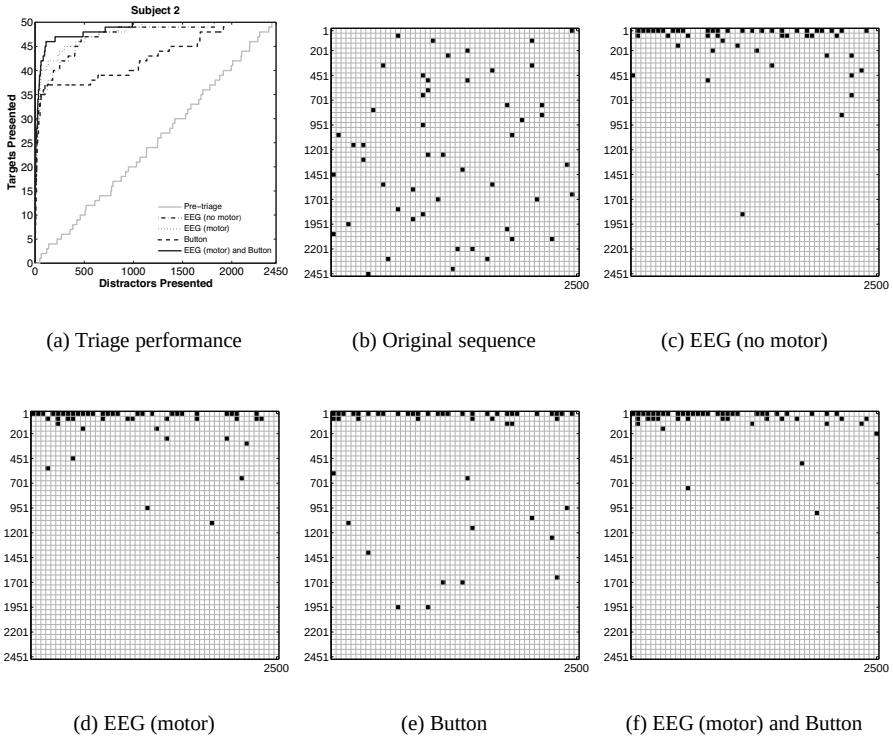
1.7 Conclusion

Invasive and non-invasive electrophysiological recordings obtained during RSVP of natural image stimuli have shed light on the speed, variability and spatio-temporal dynamics of visual processing. Recent advances in high-spatial density EEG, real-time signal processing, machine learning and human-computer interface design have enabled these basic neuroscience findings to be used in the development of systems which could support high-throughput image triage. Further basic and applied neuroscience research by our group will consider issues related to learning/priming/habituation, the effect of subject expertise, image type and category and correlated spatio-temporal structure which is prevalent in video sequences. We continue to develop ‘application level’ demonstrations which focus on intercepting neural correlates of visual discrimination and recognition events that effectively by-pass the ‘slow and noisy’ motor response loop.

1.8 Acknowledgments

This work was supported in part by grants from the the Defense Advanced Research Projects Agency, the Office of Naval Research Multidisciplinary University Research Initiative, the National Geospatial-Intelligence Agency and the National Institutes of Health.

REFERENCES



(a) Triage performance (b) Original sequence (c) EEG (no motor)

(d) EEG (motor) (e) Button (f) EEG (motor) and Button

Figure 1.8: Triage performance for subject 2. (a) Number of target images presented as a function of the number of distractor images presented. An ideal triage system will place 50 (100%) of target images before all 2450 distractor images. The light gray curve shows the original sequence. Button-based triage is shown by the dashed curve. The dash–dot curve shows EEG-based triage during the experiment without motor response. The dotted curve shows EEG-based triage during the experiment with motor response and the thick black curve shows triage based on EEG (motor) and the button response. (b–f) Rasters showing the position of non–target (white squares) and target (black squares) within the (b) original image sequence (c) EEG (no motor)–based triage sequence, (d) EEG (motor)–based triage sequence (e) button–based triage sequence and (f) Combined EEG (motor) & Button–based triage sequence. The first and last images in each sequence are shown by the squares in the upper left and lower right of each raster respectively.

1. K.H. Britten, W.T. Newsome, M.N. Shadlen, S. Celebrini, and J.A. Movshon. A relationship between behavioral choice and visual responses of neurons in macaque MT. *Visual Neuroscience*, 14:87–100, 1996.
2. K.H. Britten, M.N. Shadlen, W.T. Newsome, and J.A. Movshon. The analysis of visual motion: A comparison of neuronal and psychophysical performance. *Journal of Neuroscience*, 12(12):4745–4765, 1992.
3. M.M. Bronstein, A.M. Bronstein, M. Zibulevsky, and Y.Y. Zeevi. Blind deconvolution of images using optimal sparse representations. *IEEE Trans Image Process*, 14(6):726–736, 2005.
4. M. Cheng, X. Gao, S. Gao, and D. Xu. Design and implementation of a brain-computer interface with high transfer rates. *IEEE Trans Biomed Eng*, 49(10):1181–1186, 2002.
5. S.C. Dakin. What causes non-monotonic tuning of fMRI response to noisy images? *Current Biology*, 12:476–477, 2002.
6. A. Delorme and S. Makeig. EEG changes accompanying learned regulation of 12-hz EEG activity. *IEEE Trans Neural Syst Rehabil Eng*, 11(2):133–137, 2003.
7. A.D. Gerson, L.C. Parra, and P. Sajda. Cortical origins of response time variability during rapid discrimination of visual objects. *Neuroimage*, 28(2):342–53, 2005.
8. D.M. Green and J.A. Swets. *Signal detection theory and psychophysics*. Wiley, New York, 1966.
9. U. Hasson, I. Levy, M. Behrmann, T. Hendler, and R. Malach. Eccentricity bias as an organization principle for human high-order object areas. *Neuron*, 34:479–490, 2002.
10. H.R. Heekeren, S. Marrett, P.A. Bandettini, and L.G. Ungerleider. A general mechanism for perceptual decision-making in human brain. *Nature*, 431:859–862, 2004.
11. P. Hoel, S. Port, and C. Stone. *Introduction to statistical theory*. Houghton Mifflin, Boston, 1971.
12. D.A. Jeffreys. Evoked studies of face and object processing. *Visual Cognition*, 3:1–38, 1996.
13. J.S. Johnson and B.A. Olshausen. Timecourse of neural signatures of object recognition. *Journal of Vision*, 3:499–512, 2003.
14. T. Kammer, L. Lehr, and K. Kirschfeld. Cortical visual processing is temporally dispersed by luminance in human subjects. *Neuroscience Letters*, 263:133–136, 1999.
15. M. Kaper, P. Meinicke, U. Grossekathoefer, T. Lingner, and H. Ritter. BCI Competition 2003–Data set IIb: support vector machines for the P300 speller paradigm. *IEEE Trans Biomed Eng*, 51(6):1073–1076, 2003.
16. S.P. Kelly, E.C. Lalor, R.B. Reilly, and J.J. Foxe. Visual spatial attention tracking using high-density ssvep data for independent brain-computer communication. *IEEE Trans Neural Syst Rehabil Eng*, 13(2):172–178, 2005.
17. C. Keysers, D.-K Xiao, P. Foldiak, and D.I. Perrett. The speed of sight. *Journal of Cognitive Neuroscience*, 13(1):90–101, 2001.
18. S. Lemm, B. Blankertz, G. Curio, and K.-R. Müller. Spatio-spectral filters for improving the classification of single trial EEG. *Biomedical Engineering, IEEE Transactions on*, 52(9):1541–1548, 2005.
19. J. Liu, A. Harris, and N. Kanwisher. Stages of processing in face perception: An MEG study. *Nature Neuroscience*, 5(9):910–916, 2002.
20. A. Luo and P. Sajda. Learning discrimination trajectories in EEG sensor space: application to inferring task difficulty. *Journal of Neural Engineering*, 3:L1–L6, 2006.

21. S. Makeig, A. Delorme, M. Westerfield, J. Townsend, E. Courchense, and T. Sejnowski. Electroencephalographic brain dynamics following visual targets requiring manual responses. *PLoS Biology*, 2(6):747–762, June 2004.

22. S. Makeig, M. Westerfield, T.-P. Jung, J. Covington, J. Townsend, T. Sejnowski, and E. Courchesne. Independent components of the late positive response complex in a visual spatial attention task. *Journal of Neuroscience*, 19:2665–2680, 1999.

23. K.R. Muller, C.W. Anderson, and G.E. Birch. Linear and non-linear methods for brain-computer interfaces. *IEEE Trans. Neural Sys. Rehab. Eng.*, 11:165–169, 2003.

24. W.T. Newsome, K.H. Britten, and J.A. Movshon. Neural correlates of a perceptual decision. *Nature*, 341(6237):52–54, 1989.

25. L.C. Parra, C.D. Spence, A.D. Gerson, and P. Sajda. Recipes for the linear analysis of EEG. *Neuroimage*, 28(2):326–341, 2005.

26. G. Pfurtscheller. Functional topography during sensorimotor activation studied with event-related desynchronization mapping. *J Clin Neurophysiol*, 6:75–84, 1989.

27. G. Pfurtscheller and C. Neuper. Motor imagery and direct brain-computer communication. *Proceedings of the IEEE*, 89(7):1123 – 1134, 2001.

28. M.G. Philiastides and P. Sajda. Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex*, 16(4):509–518, 2006.

29. A. Puce, Allison T., Asgari M., J.C. Gore, and G. McCarthy. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. *Journal of Neuroscience*, 16:5205–5215, 1996.

30. R.F. Quick. A vector magnitude model of contrast detection. *Kybernetik*, 16:65–67, 1974.

31. B. Rossion, C. Joyce, G.W. Cottrell, and M.J. Tarr. Early laterization and orientation tuning for face, word, object processing in the visual cortex. *NeuroImage*, 20:1609–1624, 2003.

32. M. K. Scheffers, R. Johnson, and D. S. Ruchkin. P300 in patients with unilateral temporal lobectomies: the effects of reduced stimulus quality. *Psychophysiology*, 28:274–284, 1991.

33. S. Thorpe, D. Fize, and C. Marlot. Speed of processing in the human visual system. *Nature*, 381:520–522, 1996.

34. R. VanRullen and S.J. Thorpe. The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience*, 13(4):454–461, 2001.

35. JR Wolpaw and DJ McFarland. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proc Natl Acad Sci U S A.*, 101(51):17849–54., 2004.

36. J.R. Wolpaw, DJ McFarland, G.W. Neat, and C.A. Forneris. An eeg-based brain-computer interface for cursor control. *Electroencephalogr Clin Neurophysiol*, 78:252–9, 1991.